

Best of the Best Technology Transitions to Industry and the Fleet

ONR Naval-Industry R&D Partnership Conference

August 5, 2004

Jason Wiener

CEO

Dipsie, Inc

jason.wiener@dipsie.com



Why does the world need another search engine?

- Less than 1% of the Web has been indexed by search engines
 - Sites that utilize dynamic technologies such as cookies, databases, scripted pages and forms are missed
- Users must sift through dozens of pages to find search results
- Current technology isn't getting better. It's getting worse.
- Search is quickly becoming the third-largest Internet market
 - >550 million searches are performed per day
 - est \$3 Billion market in 2004
 - est \$7 Billion market by 2007



The way search works today

- Data Mining
 - Search Engines only crawl static HTML pages and some dynamic pages
- Information Retrieval
 - Term Frequency/Keyword Matching
 - The number of times and where a term shows up in a document compared to how it shows up in the entire collection of documents
 - Clusters
 - Similar documents are grouped based on the similarity of terms and phrases
 - Link Weighting (Authoritative Hubs)
 - Scoring based on whom links to whom and the “authoritative” perception of sites along the link line
 - Latent Semantic Analysis/Vector Space Models
 - Concepts/topics present based on key nouns and verbs within a document/site



Search's inefficiencies

- Data Mining
 - Search Engines are missing a landscape of pages estimated to be between 200 and 400x the size of current search repositories
- Information Retrieval
 - Term Frequency/Keyword Matching
 - A term/phrase occurring frequently or in high value positions within a document doesn't necessarily mean that it (or the document) is of high value to the user
 - Clusters
 - Is term centric, not "what" contextually the user is searching for
 - Link Weighting (Authoritative Hubs)
 - If a site isn't popular, it's penalized
 - Latent Semantic Analysis/Vector Space Models
 - More accurate, but inefficient when trying to scale for large collections



What does that mean for users?

- Searches leverage a less than comprehensive sample of the Web
- Lower quality results are returned to queries
- More time is wasted
- Satisfaction decreases
- Frustration increases



Dipsie – better results from more content

- Dipsie indexes more of the Web (99+%)
 - Cookies that manage sessions, tracking and load balancing
 - Page templates that deliver database-driven content
 - Forms and client-side scripting that drive user experience
 - Static HTML web pages
- Dipsie returns higher quality results using language-based ranking methods
 - Lexical energy based ranking
 - Relates the utility of terms and phrases to documents indexed
 - Contextual taxonomy generation
 - Pre categorized results based semantically on the query
 - Semantic query controls
 - Interface controls tailored to best represent query points for a given topic
 - Dipsie strives to minimize the clicks necessary to get from query to content



Search Landscape



Search evolution in the next 5 years?

- Ability to identify what users are looking for semantically
- Larger, more stable and scalable indexes
- Search access points augmenting query presentation process
- Ability to drill down and find information more quickly & easily
- Immediate action items following results use
- Smaller results sets with higher value content



What will this mean for users?

- Larger, more stable content collections to search from
- Intelligent semantic results presentation
- Less time searching, more time finding
- Higher quality of life.

